ELSEVIER

Discrete Mathematics 169 (1997) 257–262

DISCRETE MATHEMATICS

Note

Two applications of separating systems to nonadaptive procedures

Anthony J. Macula *

Department of Mathematics, State University of New York, College at Geneseo, Geneseo, NY 14454, USA

Received 25 August 1995; revised 8 November 1995

Abstract

Let $[n]$ denote $\{1,2,\ldots,n\}$. A set system σ on $[n]$ is called a *separating system* on $[n]$ if for each pair of distinct elements in $[n]$, there is an $M \in \sigma$ that contains exactly one member of that pair. A *separating system* τ on $[n]$ is called *totally separating* if for each pair of distinct elements j, j' in $[n]$, there are disjoint $M, M' \in \tau$ with $j \in M$ and $j' \in M'$. In this paper, we discuss two applications of separating systems. In the first application, we give an easy method of constructing a class of single error-correcting double error-detecting codes over an Abelian group alphabet. In the second application, we construct a class of $\bar{3}$-separable matrices. Separable matrices are important notions in the theories of nonadaptive group testing and binary superimposed codes.

Keywords: Separating systems; Error-correcting codes; $\bar{3}$-separable matrices; Nonadaptive group testing; Binary superimposed codes

1. Separating systems and encoding over an Abelian group

Let $[n]$ denote $\{1,2,\ldots,n\}$. Given set S, $|S|$ denotes its cardinality. In Sections 1 and 2, G denotes an Abelian group with operation $+$ and identity e. We write G^n to denote the direct product of G with itself n times. We let $\bar{e} = (e,\ldots,e)$ denote the identity in G^n. For $\boldsymbol{x} \in G^n$ and $j \in [n]$ let $\boldsymbol{x}_j$ be the jth component of $\boldsymbol{x}$. For $S \subset [n]$ and $\boldsymbol{x} \in G^n$, we define $S\boldsymbol{x} \in G$ by $S\boldsymbol{x} = \sum_{j \in S} \boldsymbol{x}_j$. For an ordered family $\gamma = (S_1, S_2, \ldots, S_k)$ of nonempty subsets of $[n]$, we let $\|\gamma\|$ denote the number of components of γ.

Definition 1. For an ordered family, $\gamma = (S_1, S_2, \ldots, S_k)$, of nonempty subsets of $[n]$, we define the homomorphism $h_\gamma : G^n \to G^{\|\gamma\|}$ by $h_\gamma(\boldsymbol{x}) = (S_1\boldsymbol{x}, S_2\boldsymbol{x}, \ldots, S_k\boldsymbol{x})$.We let $K_G^n(\gamma)$ denote the kernel of h_γ, and we just write $K(\gamma)$ when the context is clear.

* E-mail: macula@geneseo.edu.

PII S0012-365X(96)00091-X

For $\boldsymbol{x},\boldsymbol{y} \in G^n$, the *Hamming distance*, $d(\boldsymbol{x},\boldsymbol{y})$, is the number of entries in $\boldsymbol{x}$ and $\boldsymbol{y}$ that are different. For a subset C in G^n, let $d(C) = \min\{d(\boldsymbol{x},\boldsymbol{y})\colon\ \boldsymbol{x},\boldsymbol{y} \in C\}$. Then C is a single error-correcting double error-detecting code if and only if $d(C) = 4$. See [5].

Definition 2. An ordered family $\sigma = (S_1, S_2, \ldots, S_k)$ on $[n]$ is called an *ordered separating system* on $[n]$ if for distinct elements $j, j' \in [n]$, there is a component S_i of σ such that either $j \in S_i$ and $j' \notin S_i$, or $j \notin S_i$ and $j' \in S_i$. If S_i satisfies the above condition for $j, j' \in [n]$, then we say that S_i *separates* j from j'.

Let $[n] - S$ denote the complement of $S \subset [n]$. It is easy to verify that $\sigma = (S_1, S_2, \ldots, S_k)$ is an ordered separating system on $[n]$ if and only if for each $j \in [n]$, we have that

$$\{j\} = \bigcap_{i\in[k]} (\{S_i\colon\ j \in S_i\} \cup \{[n] - S_i\colon\ j \notin S_i\}).$$

It is folklore [2] that $\min\{\|\sigma\|\colon\ \sigma \text{ is an ordered separating system on } [n]\} = \lceil \lg n \rceil$.[1]

Proposition 1. *If σ is an ordered separating system on $[n]$ that has $[n]$ as one of its components, then for any G, we have that $d(K_G^n(\sigma)) \geqslant 4$.*

Proof. Let $\boldsymbol{x},\boldsymbol{y} \in G^n$. Without loss of generality let $\sigma = (S_1, S_2, \ldots, S_k)$ where $S_1 = [n]$. We argue by contradiction. If $d(\boldsymbol{x},\boldsymbol{y}) = 1$, then $S_1\boldsymbol{x} \neq S_1\boldsymbol{y}$. So x and y cannot both be in $K(\sigma)$. If $d(\boldsymbol{x},\boldsymbol{y}) = 2$, then there are *exactly* two distinct elements $j, j' \in [n]$ for which $\boldsymbol{x}_j \neq \boldsymbol{y}_j$ and $\boldsymbol{x}_{j'} \neq \boldsymbol{y}_{j'}$. Since there is a component S_i of σ such that either $j \in S_i$ and $j' \notin S_i$, or $j \notin S_i$ and $j' \in S_i$, it follows that $S_i\boldsymbol{x} \neq S_i\boldsymbol{y}$. So $\boldsymbol{x}$ and $\boldsymbol{y}$ cannot both be in $K(\sigma)$. If $d(\boldsymbol{x},\boldsymbol{y}) = 3$, then there are *exactly* three distinct elements $j_1, j_2, j_3 \in [n]$ for which $\boldsymbol{x}_{j_1} \neq \boldsymbol{y}_{j_1}$, $\boldsymbol{x}_{j_2} \neq \boldsymbol{y}_{j_2}$, and $\boldsymbol{x}_{j_3} \neq \boldsymbol{y}_{j_3}$. We have two cases. If there is a component S_i of σ such that $|S_i \cap \{j_1, j_2, j_3\}| = 1$, then it follows that $S_i\boldsymbol{x} \neq S_i\boldsymbol{y}$. So $\boldsymbol{x}$ and $\boldsymbol{y}$ cannot both be in $K(\sigma)$. If not, then because there is an S_i of σ that separates j_1 from j_2, it follows that $|S_i \cap \{j_1, j_2, j_3\}| = 2$. Without loss of generality, suppose $S_i \cap \{j_1, j_2, j_3\} = \{j_1, j_2\}$. Now either $S_1\boldsymbol{x} \neq S_1\boldsymbol{y}$ or $S_i\boldsymbol{x} \neq S_i\boldsymbol{y}$ because if both $S_1\boldsymbol{x} = S_1\boldsymbol{y}$ and $S_i\boldsymbol{x} = S_i\boldsymbol{y}$, then it follows that $\boldsymbol{x}_{j_3} = \boldsymbol{y}_{j_3}$. So $\boldsymbol{x}$ and $\boldsymbol{y}$ cannot both be in $K(\sigma)$. □

Corollary 1. *For any G, there is a subgroup K of G^n with $|K| \geqslant |G|^{n-\lceil \lg n \rceil - 1}$ and $d(K) \geqslant 4$.*

Proof. Since $\min\{\|\sigma\|\colon\ \sigma \text{ is an ordered separating system on } [n]\} = \lceil \lg n \rceil$, there is an ordered separating σ with $\|\sigma\| = \lceil \lg n \rceil + 1$ which has $[n]$ as one of its components. Take $K = K_G^n(\sigma)$.[2] Then $|K| \geqslant |G^n|/|G^{\|\sigma\|}| = |G|^{n-\lceil \lg n \rceil - 1}$. □

[1] By $\lg n$ we mean $\log_2 n$.

[2] If we think of a finite field F as group G, then the codes $K_F^n(\gamma)$ are linear.

2. Decoding

Let $\sigma = ([n], S_2, \ldots, S_k)$ be an ordered separating system on $[n]$. Decoding $K_G^n(\sigma)$ is very simple. Let e be the identity in G. Suppose a received transmission $\boldsymbol{x}$ has at most two errors. By Proposition 1, it follows that $\boldsymbol{x}$ is error-free if and only if $h_\sigma(\boldsymbol{x}) = \bar{e}$. So suppose that $h_\sigma(\boldsymbol{x}) \neq \bar{e}$. Consider the family of subsets $\{S_i\colon\ S_i\boldsymbol{x} \neq e\} \cup \{[n] - S_i\colon\ S_i\boldsymbol{x} = e\}$ of $[n]$. Because σ is an ordered separating system, there is at most one element in $\bigcap(\{S_i\colon\ S_i\boldsymbol{x} \neq e\} \cup \{[n] - S_i\colon\ S_i\boldsymbol{x} = e\})$. If $\bigcap(\{S_i\colon\ S_i\boldsymbol{x} \neq e\} \cup \{[n] - S_i\colon\ S_i\boldsymbol{x} = e\}) = \emptyset$, then $\boldsymbol{x}$ had two errors. If $\{j^*\} = \bigcap(\{S_i\colon\ S_i\boldsymbol{x} \neq e\} \cup \{[n] - S_i\colon\ S_i\boldsymbol{x} = e\})$, then either $\boldsymbol{x}_{j^*}$ is the single error or two errors have occurred. Since $S_1\boldsymbol{x}$ is supposed to be e and $j^* \in S_1 = [n]$, consider $g = -\sum_{j \in [n], j \neq j^*} \boldsymbol{x}_j$. It follows that if $\boldsymbol{x}_{j^*}$ is the single error, then the correct value of $\boldsymbol{x}_{j^*}$ is g. More precisely, let $\boldsymbol{x}(j^*, g) \in G^n$ be the element that results by taking $\boldsymbol{x}$ and replacing its j^* component with g. Now, because σ is an ordered separating system, it follows that $h_\sigma(\boldsymbol{x}(j^*, g)) = \bar{e}$ if and only if $\boldsymbol{x}(j^*, g)$ is the intended message.

3. Group testing

Let us suppose that we have a finite ground set containing elements which can be characterized as being either good or defective. We refer to the collection of defective elements, which is unknown at the outset, as the *defective subset*. In the abstract *group testing problem*, we are faced with identifying the defective subset by performing a series of 0, 1 tests on subsets of the ground set. A test result is 1 if a defect is present in a tested subset; the test result is 0 otherwise. In nonadaptive group testing (NGT), we have the added difficulty of deciding exactly which subsets to test before any testing occurs. Thus, the testing procedure for an NGT problem cannot be adapted to use partial information obtained from some of the other tests.[3]

We identify the ground set with $[n]$. For $k \in [n]$, let $\left[\binom{n}{k}\right]$ denote the family of subsets of $[n]$ with cardinality less than or equal to k. A family $\mu = \{M_i\}$ of distinct subsets of $[n]$ is called a *set system* on $[n]$. Any set system μ on $[n]$ can be identified with a $|\mu| \times n$ *incidence matrix* for which the jth column denotes the jth element of $[n]$, the ith row denotes the ith member, M_i, of μ, and the (ij)th entry is 1 if and only if $j \in M_i$. We identify a given set system with its incidence matrix. Conversely, if an $m \times n$ incidence matrix has distinct rows, we identify it with a set system on $[n]$ as described above. In other words, given an $m \times n$ incidence matrix with distinct rows, we think of the rows as a set system on the set of columns. With this in mind, let $\boldsymbol{r}_i(\mu)$ and $\boldsymbol{c}_j(\mu)$ be the ith row vector and jth column vector of μ respectively. For a binary vector $\boldsymbol{x}$ let x_i denote the ith component of $\boldsymbol{x}$. The *boolean sum* of two binary vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ is defined coordinatewise using the rule $x_i \vee y_i = 0$ if and only if x_i and y_i are both zero; otherwise $x_i \vee y_i = 1$.

[3] Note, every parallel algorithm is a nonadaptive algorithm.

Consider an $m \times n$ 0, 1 matrix μ with distinct rows. The matrix μ is called $\bar{d}$-*separable* if no two boolean sums of the form, $\bigvee_{j \in S} \boldsymbol{c}_j(\mu)$, with $S \in \left[\left[{n \atop d}\right]\right]$, are the same. The matrix μ is called d-*separable* if no two boolean sums of the form, $\bigvee_{j \in S} \boldsymbol{c}_j(\mu)$, with $|S| = d$, are the same. If we view μ as a set system on $[n]$, then it is easy to see that μ is $\bar{d}$-separable if and only if for any pair of distinct sets S_1 and S_2 in $\left[\left[{n \atop d}\right]\right]$, there is an $M \in \mu$ such that either $M \cap S_1 \neq \emptyset$ and $M \cap S_2 = \emptyset$, or $M \cap S_1 = \emptyset$ and $M \cap S_2 \neq \emptyset$.

Now suppose we have at most d defects in $[n]$ and we have a reliable testing procedure that will detect the presence of a defect in a tested subset. Then a $\bar{d}$-separable matrix μ on $[n]$, with $|\mu| = m$, provides the basis for an NGT algorithm that identifies the defective subset. Indeed, by testing each row of μ, we define an *output vector* $\boldsymbol{o}(\mu) = (t_1, \ldots, t_m)$, where t_i is 1 if a defect is present in $\boldsymbol{r}_i(\mu)$ and 0 if not. Now if $\boldsymbol{o}(\mu)$ is the zero vector, then $D = \emptyset$. If not, then because μ is $\bar{d}$-separable, it follows that there is a unique nonempty subset D of the $\boldsymbol{c}_j(\mu)$, with $|D| \leqslant d$, whose boolean sum is $\boldsymbol{o}(\mu)$. In this case, the defective subset is $\{j\colon\ \boldsymbol{c}_j(\mu) \in D\}$. The same argument is true if we test the rows of a d-disjunct matrix μ. However, it is computationally much easier to solve for D if the matrix is d-disjunct, for then, $D = \{\boldsymbol{c}_j(\mu)\colon\ \boldsymbol{c}_j(\mu) \vee \boldsymbol{o}(\mu) = \boldsymbol{o}(\mu)\}$. See [1].

4. A class of $\bar{3}$-separable matrices

A set system χ on $[n]$ is called a *completely separating system* on $[n]$ if for each pair of distinct elements $j, j' \in [n]$, there are M and M' in χ such that $j \in M$ and $j' \notin M$, and $j' \in M'$ and $j \notin M'$. A set system τ on $[n]$ is called a *totally separating system* on $[n]$ if for each pair of distinct elements $j, j' \in [n]$, there are disjoint $M, M' \in \tau$ with $j \in M$ and $j' \in M$. Clearly, totally separating implies completely separating. (See [2] for a nice survey of these and related topics.) Finally, a set system is a k-*cover* of $[n]$ if each element of $[n]$ is contained in at least k members of that system.

Definition 3. Given a set system μ on $[n]$, we define the new set system $\bar{\mu}$ on $[n]$ by $\bar{\mu} = \{T\colon\ T = M \cap M' \neq \emptyset \text{ and } M, M' \in \mu\}$.

The proof of the following is straightforward. A more general version of the result is in [4].

Lemma 1. *Let χ be a completely separating system on $[n]$ and let j_1, j_2, j_3 be three distinct elements in $[n]$. Then for each i with $1 \leqslant i \leqslant 3$, there is a $T_i \in \bar{\chi}$ such that $T_i \cap \{j_1, j_2, j_3\} = \{j_i\}$.*

Lemma 2. *Let τ be a totally separating 2-cover on $[n]$, then $\bar{\tau}$ is $\bar{3}$-separable on $[n]$.*

Proof. Given S_1 and S_2 in $\left[\left[{n \atop 3}\right]\right]$, we need to see that there is an $T_{12} \in \bar{\tau}$ such that either $T_{12} \cap S_1 \neq \emptyset$ and $T_{12} \cap S_2 = \emptyset$, or $T_{12} \cap S_1 = \emptyset$ and $T_{12} \cap S_2 \neq \emptyset$. If $S_1 \subset S_2$, then the existence of T_{12} follows from Lemma 1.

Now suppose that $S_1 \not\subset S_2$ and $S_2 \not\subset S_1$. Then there are j_1 and j_2 with $j_1 \in S_1$ and $j_1 \notin S_2$, and $j_2 \in S_2$ and $j_2 \notin S_1$. Since there are disjoint sets M_1 and M_2 in τ with $j_1 \in M_1$ and $j_2 \in M_2$, then $|M_1 \cap S_2| \leqslant 2$. We now check three cases.

Case 1: $|M_1 \cap S_2| = 0$. Since τ is a 2-cover, take T_{12} to be the intersection of M_1 with any other member of τ that contains j_1.

Case 2: $|M_1 \cap S_2| = 1$. Let $\{j_3\} = M_1 \cap S_2$. There is an $M_3 \in \tau$ with $j_1 \in M_3$ and $j_3 \notin M_3$. Take $T_{12} = M_1 \cap M_3$. Then $T_{12} \cap S_1 \neq \emptyset$ and $T_{12} \cap S_2 = \emptyset$.

Case 3: $|M_1 \cap S_2| = 2$. Let $j_4 \in M_1 \cap S_2$. We have two subcases:

(a) $|M_1 \cap S_1| = 1$. Then $\{j_1\} = M_1 \cap S_1$. So there is an $M_4 \in \tau$ with $j_4 \in M_4$ and $j_1 \notin M_4$. Now take $T_{12} = M_1 \cap M_4$. Then $T_{12} \cap S_1 = \emptyset$ and $T_{12} \cap S_2 \neq \emptyset$.

(b) $|M_1 \cap S_1| = 2$. Then $|M_2 \cap S_1| \leqslant 1$. Reapply Cases 1 and 2 with M_1 and M_2, and S_1 and S_2 interchanged. □

In Definition 4, we show that it is easy to construct a totally separating τ on $[n]$ with $|\tau| \leqslant r\lceil \log_r n \rceil$. The following scheme depicts the process with $r = 3$. It is left to the reader to verify that the defined set systems are indeed totally separating.

$$\begin{pmatrix} 1&0&0 \\ 0&1&0 \\ 0&0&1 \end{pmatrix} \Rightarrow \begin{pmatrix} 1&1&1&0&0&0&0&0&0 \\ 0&0&0&1&1&1&0&0&0 \\ 0&0&0&0&0&0&1&1&1 \\ 1&0&0&1&0&0&1&0&0 \\ 0&1&0&0&1&0&0&1&0 \\ 0&0&1&0&0&1&0&0&1 \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} 111&111&111&000&000&000&000&000&000 \\ 000&000&000&111&111&111&000&000&000 \\ 000&000&000&000&000&000&111&111&111 \\ 111&000&000&111&000&000&111&000&000 \\ 000&111&000&000&111&000&000&111&000 \\ 000&000&111&000&000&111&000&000&111 \\ 100&100&100&100&100&100&100&100&100 \\ 010&010&010&010&010&010&010&010&010 \\ 001&001&001&001&001&001&001&001&001 \end{pmatrix}.$$

Definition 4. Let r and k be positive integers. Let τ_r^1 be the $r \times r$ identity matrix. For $k > 1$, define τ_r^k by taking τ_r^{k-1} and, in it, replacing each 1 with a block r 1s and each 0 with a block r 0s. Then add r^{k-1} copies of the $r \times r$ identity matrix side by side under the newly formed rows. For n with $r^{k-1} < n \leqslant r^k$, we define a totally separating $\tau_r^k(n)$ on $[n]$ with $|\tau_r^k(n)| \leqslant r\lceil \log_r n \rceil = rk$ by deleting the last $3^k - n$ columns of τ_r^k.

Theorem 1. *For each n and r with $\lceil \log_r n \rceil > 1$, there is a $\bar{3}$-separable matrix with n columns and no more than $\binom{r\lceil \log_r n \rceil}{2} - \binom{r}{2}\lceil \log_r n \rceil$ rows.*

Proof. Consider $\overline{\tau_r^k(n)}$ with $k > 1$. Since $\tau_r^k(n)$ is a 2-cover, we can apply Lemma 2. It is easy to see that $\overline{\tau_r^k(n)}$ has no more than $\binom{r\lceil \log_r n \rceil}{2} - \binom{r}{2}\lceil \log_r n \rceil$ rows. □

Note, the value $r = 3$ almost always minimizes $|\overline{\tau_r^k(n)}|$. This improves a result of Kautz and Singleton [3] which says that for a given n there is a $\bar{3}$-separable matrix with no more than $\binom{3\lceil \log_2(n+1) \rceil}{2}$ rows. See [1, p. 87]. Also, from [5], $3\log_3 n \leqslant \min\{|\tau|: \tau \text{ is totally separating on } [n]\} \leqslant 3\log_3 n + 2$, so our construction with $r = 3$ is nearly optimal via the above method.

One closing observation is that since τ_r^k has constant column weight k and constant row weight r^{k-1}, it follows that $\overline{\tau_r^k(n)}$ has constant column weight $\binom{k}{2}$ and row weight upper bounded by r^{k-1}.

References

[1] D.-Z. Du and F.K. Hwang, Combinatorial Group Testing and Its Applications (World Scientific, Singapore, 1993).

[2] G.O.H. Katona, Renyi and the combinatorial search problems, Studia Sci. Math. Hungar. 26 (1991) 363–376.

[3] W. Kautz and R. Singleton, Nonrandom binary superimposed codes, IEEE Trans. Inform. Theory 10 (1964) 363–377.

[4] A.J. Macula, A simple construction of d-disjunct matrices with certain constant weights, Discrete Math. (to appear).

[5] F.J. MacWilliams and N.J.A. Sloane, The Theory of Error-Correcting Codes (North-Holland, Amsterdam, 1977).

[6] A.C. Yao, On a problem of Katona on minimal separating systems, Discrete Math. (1976) 193–199.